A Tutorial of Heavy-traffic Approximation in Queueing Theory

August 2021

1 Introduction

The purpose of this document is to introduce the concept of heavy-traffic approximation in Queueing Theory to science community in XXX. A system of large scale is usually difficult to analyze due to analytical intractability. For example, in call center, a simple model is the Erlang-A, aka, $M/M/N+M$, where the first M stands for Poisson arrivals, the second M for exponential service times, $N \geq 1$ for the number of servers and $+M$ for the exponential patience time. To calculate the steady-state distribution for number of contacts in the system, it involves the calculation of $N!$. See Section 5.1 for more detail. When N is large, which is often the case given the scale of operations in XXX, the computation of $N!$ can be very expensive. However, heavy-traffic theorems take advantage of large system scale and produce simple yet useful results. In this tutorial, we properly define the concept of system *scale* and *heavy-traffic*. (I also want to review how the scale affects system behavior.) We will introduce two types of heavy-traffic results and their application in queueing theory: (1) conventional heavy-traffic limits and (2) many-server heavy-traffic. But before we dive deep into the theoretical results, we first answer the question: how good is the heavy-traffic approximation, comparing to exact analysis?

2 Quality of Approximation

3 Conventional Heavy-traffic Limits

4 Many-server Heavy-traffic Limits

5 Appendix

We provide additional technical nodes in this Appendix.

5.1 Exact analysis of Erlang-A model

The Erlang-A model assumes exponential inter-arrival time with rate λ, exponential service(handle) rate μ, N parallel agents, and exponential patience time with mean $1/\gamma$. One can construct a simple Markov Chain and write the transition rate matrix to obtain the steady-state distribution (π) for number of contact in the system, denoted as X. For given N number of agents, let $\pi_n \equiv \mathbf{P}(X=n)$ denote the steady-state distribution

$$\pi_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0, \quad n = 0, 1, \cdots, N \\ \frac{1}{N!}\left(\frac{\lambda}{\mu}\right)^n \left(\prod_{j=N+1}^{n} \frac{\lambda}{N\mu + (j-N)\gamma}\right) \pi_0, \quad n = N+1, \cdots \end{cases} \tag{1}$$

where

$$\pi_0 = \left[1 + \sum_{j=1}^{N} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{N}\left(\frac{\lambda}{\mu}\right)^N \left(\sum_{i=N+1}^{\infty} \prod_{j=N+1}^{i} \frac{\lambda}{N\mu + (j-N)\gamma}\right)\right]^{-1}. \tag{2}$$